Technical Report No. 534

WHEN IN THE COURSE OF A LONGITUDINAL STUDY: DIFFERENT QUESTIONS AND SURPRISING ANSWERS

Linda A. Meyer
University of Illinois at Urbana-Champaign

July 1991

Center for the Study of Reading

TECHNICAL REPORTS

College of Education
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
174 Children's Research Center
51 Gerty Drive
Champaign, Illinois 61820

CENTER FOR THE STUDY OF READING

Technical Report No. 534

WHEN IN THE COURSE OF A LONGITUDINAL STUDY: DIFFERENT QUESTIONS AND SURPRISING ANSWERS

Linda A. Meyer
University of Illinois at Urbana-Champaign

July 1991

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

The work upon which this report is based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90, in part by a grant from the National Science Foundation, and by gifts from the Houghton-Mifflin and Silver Burdett & Ginn publishing companies. The publication does not necessarily reflect the views of the agencies supporting the research.

Abstract

In the course of a longitudinal study that has addressed the questions of how children learn to comprehend what they read and how they learn science concepts, many other research questions have emerged. This report first presents the context of the longitudinal study and the methodology used to conduct the work, and then focuses upon 14 different research questions that emerged in the kindergarten through second-grade work. Among the surprising answers to these questions are the following: (a) the length of the school day did not predict kindergarten children's school achievement; (b) individual teachers' behaviors were quite stable both from morning to afternoon and from year to year; (c) whole-class reading instruction produced the greatest gains in student achievement in reading; (d) teachers' instructional practices mediated children's performance on criterion-referenced measures; and (e) elementary-grade science textbooks were quite considerate to students.

WHEN IN THE COURSE OF A LONGITUDINAL STUDY: DIFFERENT QUESTIONS AND SURPRISING ANSWERS

In the course of our longitudinal study researching how children develop the ability to comprehend what they read and how they learn science concepts, numerous other questions have emerged. Although each of these questions is outside the global questions of the larger study, they are in and of themselves interesting topics to pursue. This report will focus on 14 of these questions. They are as follows:

1. Are entering ability and teacher behavior better predictors of student achievement than is length of school day?
2. Do kindergarten children exhibit metacognitive ability?
3. Are teaching behaviors within districts stable?
4. Are individual teachers' behaviors stable over time?
5. Does whole-class instruction in kindergarten produce higher student achievement in reading than does instruction in groups?
6. Does whole-class instruction continue to produce the highest student achievement in reading in first grade?
7. What is the relationship between the time teachers spend reading to their students and the time they spend teaching reading?
8. Does comprehension monitoring ability continue to develop in first grade?
9. Do at-risk students receive different instruction than do students who are not at risk?
10. Are first- and second-grade basals alike?
11. Does teaching mediate students' science performance on criterion-referenced measures?
12. Are science textbooks "considerate" to students?
13. How do second-grade students in whole-class reading instruction compare to students in traditional classrooms?
14. Do parents from different school districts respond consistently to science questionnaire items?

This report will begin by presenting background information about the longitudinal study to provide the context in which these questions arose. It will then proceed to address the specific questions. The discussion will draw information from a potpourri of areas, including the systematic analyses of basal readers and science textbooks and classroom observations.

Context

In 1982, personnel from the National Institute of Education (NIE) approached the Center for the Study of Reading with the request that Center researchers undertake a longitudinal study of how children learn to comprehend what they read. It was the position of the NIE that while there are hundreds, perhaps thousands of pieces of research that address many aspects of reading development, there has never been a comprehensive study of the relative impact of numerous influences upon a large group of children's reading comprehension development over time.

Before our research team began this study in the fall of 1983, the senior staff decided to focus ultimately on children's abilities to read science text as the primary indicator of their ability to "read to learn." Thus, we conceptualized this program of research as two longitudinal studies running in tandem and converging on reading in science as a content area by fifth grade. Our long-term goal was to produce causal models to explain why some children have learned to comprehend what they read better than others, as well as why the same children (or others) have learned more science concepts and processes. The study was planned to focus on two cohorts of children (about 325 per cohort) from kindergarten through at least fifth grade in reading and through sixth grade in science.

The Heuristic Model

The simplest way to think about the development of reading comprehension and of science knowledge is to view each as a function of the aptitudes and abilities children posses as they enter school. That is, what and how much children learn about science and reading in the early grades directly reflects their aptitude or verbal ability. A somewhat more complex view sees learning as a reflection both of children's aptitudes and of home-related factors, such as socioeconomic status and the amount and kind of stimulation provided by parents. Reading comprehension development and science knowledge development can also be seen as a function of the amount and kind of instructional activities children receive in school. In addition, students may be viewed as contributing to their own knowledge development through the reading they do, the television programs they watch, and the activities they choose.

To guide our work each year, we developed a simple heuristic model (Meyer, Linn, & Hastings, 1985b). The model reflects our belief that science knowledge and reading comprehension development are the result not only of student ability and prior opportunity to learn but also of immediate school and home support. Figure 1 illustrates the relationships among these factors.

[Insert Figure 1 about here.]

The model is composed of eight constructs: home background characteristics, student ability at the time they began school, the characteristics of instructional materials used to teach science, teachers' management style, teachers' instructional style, home support for science knowledge development, student ability at the end of each year, and independent reading.

The following discussion explains how we conceptualized each construct.

1. **Home background.** This construct represents the variables of parental occupation and education, the number of adults in the home, the number of older and younger siblings, and the number of hours each parent works outside the home each week.

2. **Ability, 0.** This construct represents children's verbal abilities upon entering school that are most likely to affect their reading ability or science knowledge at the end of kindergarten. Ability 0 on the model represents the children's abilities at the first testing in the fall.

3. **Materials.** This construct represents the characteristics of instructional materials that may contribute to children's reading or science development. Specifically, it represents the variables of textbook content, use, and "considerateness"--that is, the number of problems in the way science textbooks or basal readers present material.

4. **Management style.** Because we believed that teachers do not necessarily manage their classrooms in the same ways they instruct, we separated teaching initially into two constructs, management style and instructional style. Management style captures teachers' strategies for molding students' general behavior. It is composed of five classroom characteristics: (a) the amount of time teachers allocated to reading or science instruction; (b) their general praise statements to individual students; (c) their general praise statements to groups of students, such as, "Everyone is working very nicely"; (d) their critical statements directed to individual students, such as "Johnny, sit down and start to work now"; and (e) their critical statements directed to groups of students.

5. **Instructional style.** This is the second construct representing teaching. The variables of the instructional style construct are extensions of characteristics reported in research on general teaching effectiveness in other areas of elementary education such as reading and math that have demonstrated the effects that instruction can produce in these areas. Six additional classroom process variables compose this construct. Half of these variables are measures of the kinds of interactions teachers initiated with individual students or entire classes. The remaining half of these variables capture teachers' responses to students who have made errors or who can not come up with an answer. Instructional style was also characterized by the kinds of feedback--sustained, terminating, or confirming --that teachers give.

6. **Home support for science knowledge or reading comprehension development.** This construct contains three clusters of variables for science knowledge development: (a) a child's involvement in science processes with parents, (b) the frequency with which parents provide activities for their children, and (c) the prevalence of science-related books and magazines in the home to which the child has direct access. It contains up to five clusters of variables for reading comprehension development: (a) the amount the child is read to, (b) the amount the child reads, (c) resources parents' provide, (d) the frequency with which parents provide activities for their children, and (e) the prevalence of all kinds of books and magazines in the home.

7. **Independent reading.** This construct represents reading initiated by the child. We anticipated that activities such as independent reading might influence children's reading comprehension development and science knowledge in later grades, but probably not in the kindergarten through second-grade years.

8. **Ability, 1.** This construct represents students' reading ability or science knowledge in the spring of each school year.

How the Heuristic Model Has Been Actualized

Figure 2 shows annual data collection efforts for the model. Our goal has been to administer a battery of instruments twice each year to reflect changes in children's reading comprehension development and science concept and process acquisition. Most of these measures were administered individually during the children's kindergarten through second-grade years. Some individual assessment is anticipated to continue for the duration of the study.

For analyses focusing on a single year, first or second grade for example, each fall can be considered a new Time 0. On the other hand, for analyses that occur over several years, the fall first-grade test scores represent the third wave of data collection for individual students, the spring first grade as the fourth wave, and so on. At both fall and spring each year, multiple measures, norm-referenced and criterion-referenced, are administered either to individuals or groups as appropriate. Many of these are measures that were either developed for this study or were developed by others and then modified on the basis of previous research findings.

[Insert Figure 2 about here.]

Ability, 0. Thus far, many measures of student ability have been administered each fall and again in the spring. The decoding subtest of the Wide Range Achievement Test (WRAT) (Jastak, Bijou, & Jastak, 1965) is one of these instruments. It has been administered each fall to each child since the first week in kindergarten. Other fall measures include a test of the children's knowledge of environmental print, listening comprehension, passage reading comprehension, and metacognition as measured on various error detection instruments.

Classroom management. Classroom management characteristics have been derived from nine full-day observations of each teacher. These observations are also audiotaped. For this construct, the observational data include time in activities, teachers' overall grouping patterns to deliver instruction, and their praise and corrective statements to individuals or groups of children.

Instructional materials. Reading and science textbooks are analyzed each year to reveal most simply what the content is. For example, analyses of the science textbooks include the number of specific content domains and the vocabulary they present for students to learn. We also study the activities teachers are to present. Furthermore, in science textbooks, we study the ratio of lecture/discussion activities to hands-on activities, how many times teachers are to demonstrate experiments, and how often students are to complete experiments on their own. Finally, we study how "considerate" the text children read is. These are quantitative measures of text characteristics of instructional programs that others, Anderson and Armbruster (1984), for example, have determined affect students' abilities to comprehend what they read. For all of these materials' characteristics, we have analyzed every page of all science materials used in the districts participating in this study.

Teachers in District A used the Merrill (Sund, Adams, & Hackett, 1980) Accent on Science series. District B adopted Holt Science (Abruscato, Forsaceca, Hassard, & Peck, 1980), after using the McGraw Hill (Holmes, Leake, & Shaw, 1974) materials. District C teachers implemented Silver Burdett Science (Mallinson, Mallinson, Smallwood, & Valentino, 1985).

We analyzed the reading materials to learn about their contents, pedagogy, and the comprehensibility of their stories. In addition, we tracked the amount of content individual children covered. Teachers in District A used the Houghton Mifflin (Durr, LePere, Alsin, Burryon, & Shaw, 1979) reading series, whereas those in District B used the Harcourt Brace Jovanovich (Early, Cooper, & Santeusanio, 1983a, 1983b) series. Teachers in District C used Ginn's 1976 edition (Clymer, Wong, & Benedict) as their main reading program for all students. They provided additional instruction for their low-stanine children in the Distar Reading Mastery materials (Engelmann & Bruner, 1983) published by Science Research Associates (SRA).

Instructional style. The data on instruction comes from the nine full days of classroom observations of each teacher, with the exception of the Cohort 2 second-grade teachers, who were observed seven times. These data have been coded to address the four major areas of instructional characteristics depicted in Figure 2. Each instructional interaction a teacher initiates was coded, with the codes reflecting the tasks the interactions actually call upon the children to perform. For example, if a teacher says, "Open your workbooks to page 71," the children are simply to complete the procedure. If, on the other hand, a teacher says, "Everyone read aloud the words on page 71," the children are then to perform a series of word-reading tasks. The major categories into which interactions fell are (a) procedural; (b) text-tied comprehension; (c) scriptal; (d) decoding; (e) word, sentence, or paragraph reading; or (f) inferential reasoning. Text-tied interactions require students to give an answer found in the text they have read (Pearson & Johnson, 1978). These questions may be answered either explicitly or implicitly in the text. Scriptal (or background knowledge) questions have students answer from what they already know about a topic. In addition, teachers give students turns in oral or silent reading. Furthermore, especially in science instruction, teachers often ask students to formulate hypotheses, make observations regarding manipulated materials, apply a concept, or predict an outcome. Each interaction was also coded as to whether the teacher directed it to the whole class, a small group, or an individual child. Each individual and group was numbered so that analyses could focus on the type or frequency of interactions occurring at various levels (child, group, or entire class).

Feedback was coded first into 14 categories that were then collapsed generally into three major types similar to those developed by Anderson, Evertson, and Brophy (1979): Feedback in which a teacher sustains an interaction with a student by leading or giving a hint; feedback that terminates an interaction by calling on another child or ignoring the first student's mistake altogether; or feedback that simply confirms the student's response.

Ability, 1. Student ability measures for spring were chosen to reflect changes in students' development during the year. For example, we have administered the Error Detection Test (Meyer, Hastings, Linn, & Greer, 1985) to students since their kindergarten year. In addition, we have developed our own science measures for use at the end of second and third grades to reflect the content we expect students to have covered on the basis of our analyses of their textbooks during those grades. In addition, we have administered some measures such as the WRAT decoding subtest and the Woodcock Reading Mastery Tests comprehension passages (Woodcock, 1973) at least once a year (and often twice a year) since the beginning of the study.

Independent reading. We have attempted to assess students' emerging interests in various topics as well as the amount of their independent reading beginning in third grade. These procedures and results will be reported in depth with the third- through sixth-grade findings and will therefore be described only briefly here as part of the complete heuristic model. Each third grader has a notebook for recording the title and author of each book read. The students also indicate whether they actually read the book. If they finished the book, they also write briefly about how much they liked it. In addition, they write a short phrase or statement on what the book was about. The form concludes by having the students complete this phrase, "I want to learn more about _________." At some point during each round of observations, we also record the books each third-grade student has currently checked out of the school library. This serves as an additional measure of independent reading.

Home influences. Home influences have been surveyed each spring with separate questionnaires sent to parents asking about their general support for literacy-related activities. These questionnaires focus on activities such as reading to children and the frequency of trips to the library. A second questionnaire is more directed to science development. It focuses on such things as subscriptions to science magazines. The return rate for these questionnaires has always been above 84%. It approached 90% as the Cohort 1 children completed third grade and has remained at that point. Some of these influences are tapped with direct questions such as the number of hours of TV that children watch each day, while other constructs are measured by giving the parents a paragraph describing a process-oriented scene, such as cooking with their children, and asking them to respond to a number of statements as if they were in that setting.

The Setting

Five school districts were asked to make six-year commitments to participate in this research. Each district was chosen for several reasons. First, we were interested in districts with reputations for fairly stable student populations. Second, we wanted districts with reputations for overall high student achievement. Third, when possible, we wanted to work with all children at the appropriate grade levels in the district, thereby reducing the potential for problems from attrition as families move within their communities. Fourth, we wanted school districts that were reasonably close to our university community. Fifth, we needed districts that would grant permission at the school or superintendent level to allow a study of this duration so that we could guarantee that all teachers would participate. Three districts agreed to participate fully.

District A. District A has a reputation for high student performance in reading and average performance in science. This is a farming/small-town community that is somewhat isolated. It has one elementary school. There are approximately 80 children per cohort in this setting, and all children in the district at the designated grade levels participate in the study.

District B. District B has a reputation for average student performance in reading and average performance in science. This is a "commuter" district for many families, who travel about 20 minutes to work. All children in the district at the appropriate grade levels participate in this study. There are approximately 150 children in each cohort in District B. They attend one grade school.

District C. District C is often described as a "microcosm of the universe." The district has the racial and ethnic mixture of many inner-city schools, yet its student population is far more stable. The district has a reputation for high student performance in reading. One elementary school from the district participates in this study. There are about 80 children per cohort.

Pilot school. One school from a fourth district also participates in this study as our designated pilot school. Each year, all of the new measures are piloted in this school and then subsequently revised. The student performance in this school tends to be just about normally distributed, and there is a multi-ethnic student population. About 50 students and parents at the same grade level as Cohort 1 children participate in the study from this school.

There were no significant differences in student performance between these districts on the decoding subtest of the WRAT as students entered kindergarten (Linn & Meyer, 1985).

Selected Questions

This section presents interesting questions and surprising findings from our study. The results are surprising either because they contradict the work of other researchers or because they are counter intuitive. These results, presented by grade level, are from analyses completed on two cohorts of children, approximately 625 students, assigned to about 40 teachers.

Questions Emerging from Kindergarten Findings

1. Are entering ability and teacher behavior better predictors of student achievement than is length of school day? In our sample, two districts have half-day kindergarten classes and one district has whole-day classes. One would expect whole-day kindergarten programs to yield higher student achievement than half-day programs because of the greater opportunities for instruction. Several studies (Jarvis & Molnar, 1983; Johnson, 1974; Oliver, 1980; Winter & Klein, 1970) have debated the merits of whole-day versus half-day programs without studying the academic effects. Surprisingly, we found that children in whole-day classes did not have higher student achievement at the end of kindergarten than the children in half-day programs. In fact, on average, students who had been in the whole-day programs were the lowest performers on several measures of achievement.

Of all instruments we administered, including measures of environmental print, word recognition, listening comprehension, oral language ability, and general science knowledge, we will focus upon results for the Chicago Reading Test (Barr, 1983) because as a test of letter sounds, word endings, word families, and random words, it proved most sensitive to differences in instructional emphasis in kindergarten programs. In a set of regression analyses, the classroom process variables that were associated significantly with performance on the Chicago were total teacher-directed interactions (decoding and comprehension); sustained feedback, procedures teachers use to maintain interactions with a child who has made an error until she or he can produce a correct response; and time spent in reading instruction. These variables represent characteristics that collectively suggest that the more businesslike, structured kindergarten environment results in higher student achievement in reading at that grade level (Meyer, Hastings, Wardrop, & Linn, 1988). Therefore, length of school day failed to predict student achievement whereas numerous teacher behaviors and entering ability were predictors of achievement.

2. Do kindergarten children exhibit metacognitive ability? It is commonly believed (Markman, 1977, 1979) that children develop the ability to monitor their own comprehension during adolescence. The highest performing children at the end of kindergarten (those children who were actually reading), performed well on "metacognitive" (error detection) silent reading tasks. Evidence of these 6-year-olds' ability to monitor their comprehension was clear on instructional tasks even before the spring of their kindergarten year. For this reason, children who scored 10 or above on the Woodcock reading comprehension passages subtest also received the Weber Comprehension Test (Weber, 1971), an instrument actually designed for use with inner-city children at the end of third grade as a test of reading comprehension. To complete the tasks on this test correctly, students had to select a word that "spoiled the meaning" from a passage composed of up to three short sentences. End-of-kindergarten high performers could not only perform these tasks, but they often laughed and made comments about the passages, thereby giving clear evidence that they understood what they were reading (Meyer, Hastings, Linn, & Greer, 1986). These surprising findings suggest that metacognitive ability may develop in tandem with word-recognition and word-meaning ability.

3. Are teaching behaviors within districts stable? One might expect great individual differences among teachers in the same school districts. We found, however, that the behaviors of teachers in District A tended to be more similar to each other than to the behaviors of teachers in Districts B or C. Teachers' instructional characteristics within districts were very similar when charting their time or frequencies of interactions within literacy-related or science activities. These results were stable from morning to afternoon for teachers who taught half-day classes (Linn & Meyer, 1985; Meyer, Linn, & Hastings, 1985a; Meyer, Linn, & Hastings, in press).

4. Are individual teachers' behaviors stable over time? Previous research on teacher stability (Brophy, 1972; Rogosa, Floden, & Willett, 1984; Rosenshine, 1970; Shavelson & Dempsy-Atwood, 1976) suggests that teachers' behaviors are quite unstable. However, we found teachers' time in reading and their frequencies of interactions during reading instruction were very stable from morning to afternoon and from year to year (Meyer, Linn, & Hastings, in press). Figures 3, 4, and 5 illustrate these results.

[Insert Figure 3 about here.]

Figure 3 shows plots of two kindergarten teachers from District A. The X axis represents nine rounds of full-day observations while the Y axis shows minutes spent in decoding during each round. Despite variance from round to round, the teachers are very stable in the time allocated to decoding when comparing their morning and afternoon classes.

Figure 4 shows the District B teachers on these same measures.

[Insert Figure 4 about here.]

Comparable results of stability are found when plotting the frequency of these same teachers' interactions. Figure 5 shows these tallies for the District A teachers.

[Insert Figure 5 about here.]

This figure shows the number of times each teacher presented an instructional interaction coded as decoding during each of the same nine rounds shown in the two previous figures with their morning and afternoon classes.

We believe that these data show dramatically that the teachers' behaviors were quite stable when observed with two different classes on the same days. This finding seems to contradict the earlier conclusions of Shavelson and Dempsey-Atwood (1976), Rosenshine (1970), and others such as Brophy (1972) and Rogosa et al. (1984), who reported teacher behaviors to be quite unstable. As can be seen in the figures, there is considerable variability for any given teacher from one round of observations to another, but on a given day, the scores for a single teacher with two different classes are quite comparable. This suggests that day-to-day fluctuations may well be planned and not simply random variation.

To the degree that variations over time reflect intentional changes in instruction, it seems more appropriate to study teachers at comparable points in their curricula with different classes than it is to compare differences in teachers between observational rounds. The study of kindergarten teachers afforded a unique opportunity to address this question because all but three of the teachers taught half-day classes each day (Meyer, Hastings, Wardrop, & Linn, 1988; Meyer, Linn, & Hastings, 1985a; Meyer, Linn, & Hastings, 1985b). Comparisons of whole-day teachers over years with two cohorts of children also yielded similar results for teacher stability.

This evidence of teacher stability is particularly important because researchers have often used observational data to try to explain differences in student performance, and the reliability and validity of these findings rest upon the assumption that teachers' behaviors are stable and therefore predictable. Given the apparently deliberate changes that teachers appear to make in their instructional routines over time, it may be important to obtain several observations at different intervals during the school year. In addition, full-day observations may be particularly important when studying the lower elementary grades because most lessons last for short periods of time.

5. Does whole-class instruction produce higher student achievement in reading than does instruction in groups? Teachers typically divide students into three or more groups to teach reading. Therefore, there is a very strong tradition to support the existence of reading groups. Whole-class instruction in reading, on the other hand, is virtually unheard of. Therefore, it was very surprising to find whole-class reading instruction in one of our districts. It was startling to see the results of this instruction. Whole-class instruction in reading with teachers' interactions focused on their lowest performers produced the greatest differences in achievement between classes and the least variance within classes (Linn & Meyer, 1985; Meyer, Hastings, Wardrop, & Linn, 1988; Meyer, Linn, & Hastings, 1985a). District A teachers presented their initial reading instruction in the *Alpha K Time* (Reiss & Friedman, 1976) program to whole classes. Teachers in District B grouped for instruction in Harcourt Brace Jovanovich's *Look, Listen, and Learn* (Early et al., 1983a) and in *Sounds, Symbols and Sense* (Early et al., 1983b). In contrast, District C teachers taught more informally throughout their school days. In these classes, instruction was often delivered to individual children or small groups that self-selected to work with their teachers on language experience stories or similar activities. The overall highest achievement gains across measures were found in District A (Meyer, Hastings, Wardrop, & Linn, 1988).

Questions Emerging from First-Grade Findings

1. Does whole-class instruction continue to produce the highest student achievement in reading in first grade? Were the results and patterns of whole-class instruction in kindergarten an aberration? Reading groups are the norm for reading instruction in American first grades. What would the results of whole-class instruction in reading look like in first grade? District A first-grade teachers delivered instruction in reading to entire classes just as their kindergarten colleagues had. This district's policy is to teach to entire classes since all students are expected to "cover" the same content. Teachers occasionally work with a few lower performing students grouped with some high-performing students, but these groupings are temporary. In these situations, group membership changes often so that students do not readily identify as members of a lower performing group (Meyer, Linn, & Hastings, 1987). District A students continued to outperform children from the other two districts on various reading measures at the end of first grade.

2. What is the relationship between the time teachers spend reading to their students and the time they spend teaching reading? It has been recommended by diverse groups of reading researchers and specialists that teachers spend regular classroom time reading to students (e.g., Feitelson, Kita, & Goldstein, 1986; Hewison & Tizard, 1980). This is a time-honored, almost sacred, tradition. However, we found that teachers' reading to students was negatively associated with student achievement in reading, just as it had been in kindergarten. All other large process-product studies (Brophy & Evertson, 1976; Stallings, Cory, Fairweather, & Needles, 1977; Stallings & Kaskowitz, 1974) have also found this negative relationship between teachers' reading and student achievement, despite consistently favorable results in intervention studies for reading to children (e.g., Anderson, Hiebert, Scott, & Wilkinson, 1985; Feitelson et al., 1986; Hewison & Tizard, 1980; McCormick & Mason, 1986; Tizard, Schofield, & Hewison, 1982). Within the context of this longitudinal study we hope to explain these counter-intuitive and puzzling findings.

Table 1 shows results from full-day observations of 42 kindergarten and first-grade classes. This table shows the average time (in minutes) these teachers spent in reading instruction, their total time in all instruction minus adult reading, and their story reading and decoding minutes. Time in this table represents the average minutes in each activity from the nine full days of observation each year by district.

Of most interest in Table 1 are the comparisons of minutes in reading instruction and in story reading. District A teachers (these are the teachers that have produced the overall highest reading performance) average roughly 35-40 minutes of reading instruction in kindergarten and over 40 minutes per day of reading instruction in their first-grade classes. District B and District C teachers are much more similar to each other and quite different from District A teachers at the kindergarten level. Each of these districts has only 7-8 minutes (District B) or just over 5 minutes (District C) of reading instruction per day in kindergarten. In both districts, reading instruction time increases greatly in first grade, however. It is important to keep in mind that District A teachers deliver their instruction to whole classes while District B and District C teachers divide students homogeneously into small groups for instruction. Therefore, even in first grade, District A students actually receive much more instructional time delivered by their teachers than do students in either District B or District C (Meyer, Linn, & Hastings, 1986).

[Insert Table 1 about here.]

Time spent in story reading in these three districts presents an interesting contrast to time spent in reading instruction. District C teachers spend the most time reading stories to their classes in kindergarten, almost four times the number of minutes that they spend in reading instruction. The pattern between districts is quite different for the first grades. District C first-grade teachers spend the least amount of time reading stories and District B teachers average the most minutes in this activity.

These findings allow us to make a rather simple statement to explain why kindergarten teachers' reading to students correlates negatively with those students' reading achievement. It seems that in naturalistic settings, reading to students may be thought of as an alternative to activities that result in more direct instruction in reading. It is fairly clear that teachers make choices about how they spend their time. As reading instructional time goes up in these districts, story reading time usually goes down. What remains to be seen in this longitudinal study is if time teachers spent reading at the kindergarten level pays off later in children's vocabulary growth or overall reading comprehension in later grades.

3. Does comprehension monitoring ability continue to develop? Developmental reading researchers (Markman, 1977, 1979) have assumed that children develop the ability to monitor their comprehension as they become adolescents. Contrary to these beliefs, we found some evidence that high-performing children could monitor their comprehension in kindergarten. How would all of the first graders perform on an instrument measuring their cognitive ability? Is metacognitive ability at this grade level limited to the highest performances? The kindergarten results on the Weber led to the development of our Error Detection Test (Meyer, Hastings, Linn, & Greer, 1985) to use with all children in this study at the end of their first-grade year. This instrument was developed from reading vocabulary that had appeared in the reading or science programs that all children had been taught. The first subtest, Absurd Target Words, is similar in construction to the Weber because each item was developed from several short sentences. Students were to identify a word that "spoiled the meaning." In the second subtest, Impossible Sequence, students had to identify what happened at the wrong time. Here, students read aloud to an examiner and the examiner corrected the students' oral reading errors. In addition, children were asked to support their answers. First graders were clearly able to identify both absurd target words and impossible sequences and support their answers. They averaged getting 9 out of 10 absurd target words correct and 4 out of the 6 impossible sequences. This instrument and the other new measures developed for use in this study are described with their results in Meyer, Hastings, and Linn (1988).

4. Do at-risk students receive different instruction than do students who are not at risk? Lore strongly suggests that Chapter 1 work with students begins rather uniformly in first grade. We found, however, quite different procedures and philosophies in the three districts. In first grade, the three districts began very different treatments for children at risk. In District A, for example, classroom teachers took most of the responsibility for reading instruction for their first graders. Once again, the district philosophy seemed to be that every child is to cover the district reading curriculum, so teachers set about teaching in ways to accomplish this goal. While teaching entire classes, these teachers focused on lower performers by giving them large numbers of individual turns and sustained feedback. These strategies allowed the lower performers to maintain the pace of the other students in the class. These teachers also gave large numbers of group (whole-class) turns. Students in District A chosen to receive special services were primarily those who were fairly clearly targeted for special education or children already repeating first grade.

In District B, classroom teachers also assumed most of the responsibility for reading instruction for all students in first grade. These teachers formed up to four reading groups to teach reading, just as the kindergarten teachers in this same district had done. Therefore, this district's grouping policies present a dramatic contrast to the whole-class instruction practices in District A. As in District A, few children in first grade in District B received any type of special services.

In District C, children designated as the "low-stanine" children on the basis of their kindergarten achievement test scores received special reading instruction from a teacher designated to teach those (Chapter 1) children from the onset of first grade. In this setting, the first graders who received help from the special teacher were also designated the "intensive readers." This meant that these students received reading instruction from both the low-stanine group teacher and their regular classroom teacher. The reading materials were quite different in these two instructional settings. SRA Distar Reading Mastery (Engelmann & Bruner, 1983) was used by the Chapter 1 teacher, and Ginn 720 (Clymer et al., 1976) was used in the regular classroom. The combined effectiveness of these programs for students at risk remains to be seen. Chen's (1990) study of the pooled results of these efforts revealed that by the end of second grade, successful remediation had been accomplished.

Questions Emerging from Second-Grade Findings

1. Are first- and second-grade basals alike? Basal reading programs are typically analyzed by looking at several publishers at the same grade levels (Beck & McCaslin, 1978; Chall, 1967) or by studying a variety of programs to find specific examples of activities (Osborn, 1981). Little, if any work has been done on the continuity within series, and no one has previously studied stories within series in a systematic way. Therefore, whole series typically get labeled "meaning emphasis" or "word recognition emphasis." Are these labels appropriate? Is there continuity from one level of a series to another? The contents and characteristics of the basal readers used in the three districts vary substantially. As mentioned in the description of the heuristic model, the systematic analysis of textbooks is an important part of our longitudinal study. Just as classroom observations provide information on how teachers present information to students, materials analysis results depict what students have actually covered and the quality of the text they read during instruction. Analyses of the second-grade basals used in the three districts revealed substantial differences between instructional programs. These results are summarized in Tables 2 and 3 and reported fully in Meyer, Greer, Crummey, and Boyer (in press).

[Insert Tables 2 and 3 about here.]

These analyses focused on the frequency of decoding and comprehension activities in the materials, the number of reading vocabulary words presented, and a measure of "considerateness" of text for selected matched and unmatched stories. The tallies for decoding and comprehension activities are simple counts of directives teachers are to give from every page of the teachers' guides. Words in connected text are the total number of words students read from their materials.

Story considerateness was measured by applying Beck, McKeown, Omanson, and Pople's (1984) procedures developed to modify a second-grade basal story for an experimental study to measure students' abilities to comprehend original basal and revised (more considerate) text. A summary of these results appears in Table 4. While words per incoherence and thought units per incoherence are not accurate representations of how the incoherences would be distributed throughout a story, they do give a sense of how frequently incoherences occur. We believe that it is this sense of the density of incoherences that is an important issue in beginning reading programs.

[Insert Table 4 about here.]

If the matched stories in each category are taken as a group, the stories that present a dilemma are the most comprehensible, the expository text stories the next most comprehensible, and the personification stories the least comprehensible stories in the four series. The comparisons between publishers for matched and unmatched stories reveal that the matched Ginn stories are the most comprehensible, followed fairly closely by the SRA Distar Reading Mastery stories. The Houghton Mifflin (Durr et al., 1979) stories place third, and the Harcourt Brace Jovanovich stories were found to be the least comprehensible. For unmatched stories, the SRA Distar Reading Mastery stories were the most comprehensible, followed by Houghton Mifflin, Ginn, and Harcourt Brace Jovanovich.

2. Does teaching mediate students' science performance on criterion-referenced measures? Prior to this study, there has been no long-term, systematic research on the effects of teaching on children's achievement in science. The pervasive belief is that children learn science best by doing hands-on activities (Bredderman, 1983) with little teacher direction or textbook emphasis. Therefore, despite our desire to develop tests sensitive to the content of the textbooks used in the schools participating in this study, we anticipated that performance on these instruments might not be mediated by the teaching observed in the classrooms. Nonetheless, we developed three group-administered science tests for second graders (Hastings, Meyer, & Linn, 1987). The content of these three tests was drawn from the common content domains identified in the basal science textbooks (Meyer, Crummey, & Greer, 1988) used by the three districts. Plants is a content domain about which school-age students probably have background knowledge and then learn additional information in school over several years. In the content domain Three Forms of Matter, second graders may have some background knowledge, but the domain has not yet been formally introduced in school. We also developed items for Motion as a content domain because there is a great deal of evidence (Nickerson, 1985) that students misunderstand the basic concepts of motion and because this domain has not been formally introduced in science textbooks at the second-grade level. This domain is not introduced until at least fourth grade in textbooks used in any of the districts participating in this study. All of these criterion-referenced tests were read to students.

The most interesting aspect of the results from these three tests and from the Sequential Test of Educational Progress (STEP) (Educational Testing Service, 1979) test of general science knowledge is that District B students consistently achieve the highest scores. This is the district in which teachers allocate the most time to science instruction with the greatest number of interactions and feedback to students (Linn, Meyer, & Hastings, 1987). Subsequently, they cover more science content than teachers in the other two districts.

[Insert Table 5 about here.]

Results of regression analyses that began with classroom process variables revealed that text-tied interactions and sustained feedback produced significant changes in students' achievement on the Plants test (Meyer, Hastings, & Wardrop, 1988). These results are particularly encouraging because they illustrate that teachers' instructional practices can have effects on students' achievement in science, even in the lower elementary grades.

A two-factor analysis on the three criterion-referenced tests, two norm-referenced science tests, and two norm-referenced reading tests was performed. All of the criterion-referenced instruments and the Tests of Basic Experience (TOBE 2) (CTB/McGraw-Hill, 1978), a test of general science knowledge administered in the spring of kindergarten, loaded on the first factor. The STEP and various reading instruments loaded on the second factor (Hastings, Meyer, & Linn, 1987), suggesting that the first factor may represent science concept knowledge while the second factor represents general ability. These results are particularly encouraging because they were achieved with 7-year-olds. They suggest that similar, and perhaps even more powerful results may be found with older students who receive more science instruction.

3. Are science textbooks "considerate" to students? Work by Anderson and Armbruster (1984) asserts that content area textbooks used in elementary schools today are "inconsiderate" to students because they are poorly written and illustrated. Therefore, we were quite surprised to find generally opposite results as well as substantial variance between science textbook series when we completed a quantitative analysis of several programs. In analyzing these materials, our primary questions were "What is in these books?", "How is the content presented?", and "How considerate is the text?" (see Meyer, Crummey, & Greer, 1988, for the comprehensive results of this work).

While the Holt (Abruscato et al., 1980), McGraw-Hill (Holmes et al., 1974), and Silver Burdett (Mallinson et al., 1985) programs implemented at different times in District B were found to differ substantially in the number of vocabulary words and the number of content domains presented each year, they are surprisingly alike in their overall considerateness. We used the Anderson and Armbruster (1984) categories formulated for assessing considerate text. We organized their categories under the headings of structure, content, or pictures and diagrams for this analysis. We compared common content domains at comparable grade levels using detailed definitions for each of the 10 categories. As Tables 6 and 7 illustrate, we found very few examples of the failings that Anderson and Armbruster assert characterize elementary school expository text.

These results are from reading every word and looking at every picture and diagram in six content domains (27 chapters) in five grade levels, published by four different major science textbook publishers. In Table 6, 44 of the 80 entries (55%) in the categories are zeros. The largest numbers of problems found in the texts fall clearly under the headings of incomplete background knowledge or pictures and diagrams. Table 7 shows results for the content domains of the human body, plants, the solar system, and weather. In this table, 89 of the 190 entries, or about 47% of the categories, are again zeros. The largest numbers of inconsiderate text examples fall once again under the incomplete background knowledge and hard to see or unclear pictures categories. We question the Anderson and Armbruster (1984) suggestion that all unlabeled pictures or diagrams should be counted as inconsiderate because many of the texts' pictures and diagrams in these series seemed clear and therefore without real need of labels.

[Insert Tables 6 and 7 about here.]

4. How do students in whole-class reading instruction compare to students in traditional classrooms? We continued to be surprised to find whole-class reading instruction at the second-grade level in one school, and we were particularly surprised to see how children from these classes compared in reading achievement to students from classrooms where teachers taught reading in groups. Data collected with multiple measures, group and individually administered tests of reading comprehension, show district performance fairly consistent with the kindergarten and first-grade results at the second-grade level. District A continued to have the highest student performance in reading, District B placed second, and District C held third place on three measures of reading, the WRAT, the CIRCUS Reading (Educational Testing Service, 1976), and the Error Detection Test. The instruments on which district performance shifted were the Degrees of Reading Power (DRP) (College Board, 1979) test, the Woodcock, Eugene (Engelmann & Meyer, 1974), and the Weber. On these instruments, District B students performed slightly higher than District A students. Second-grade District C students were consistently in third place except for their performance on the DRP where they were in second place.

These results are particularly interesting because they suggest that while there may be continuing positive effects from District A's kindergarten reading program and/or those teachers' whole-class reading instruction from kindergarten through second grade for decoding, other factors may be influencing comprehension development. Further analyses will address these questions to determine what mediated the children's achievement in each area of development at these grade levels.

5. Do parents from different school districts respond consistently to science questionnaire items? As described earlier, three districts participated in this study. The districts were in part selected because they represent several different types of communities, and consequently, public school settings. Therefore, one would anticipate that those differences might be reflected in parents' responses on the questionnaires we gave them. However, indices developed to analyze responses from the science questionnaires administered to parents at kindergarten, first-, and second-grade levels reveal minor differences between districts (Meyer, Linn, & Hastings, 1987). These results hold for the process involvement parents report having with their children, the number of experiences they provide for their children, and the books and other resources they provide that may be related to children's achievement in science. These results are particularly interesting given the rather diverse settings that the three districts represent. Future analyses will examine the relative impact of classroom instruction, science textbooks, and home influences on children's science concept acquisition and process knowledge.

Questions for Future Research

None of the areas of research suggested by our questions was part of the specific questions guiding our longitudinal study, yet each of them has provided an area to investigate that has resulted in findings that are important in themselves. They provide support for the importance of research that takes place in naturalistic settings where researchers begin with very general questions such as "How are these textbooks similar?" or "How do the methods of teaching reading in the lower elementary grades differ, and which method produces the best results?" This open-ended process led to findings that might simply have been overlooked if the methods had been different.

While we consider these 14 findings among the most interesting preliminary results from the first four years of our longitudinal investigation of children's reading comprehension development and science concept acquisition, many more analyses are needed to determine how these results are related. Causal models will illustrate how these outcomes work together to compose a model of how students learned to comprehend what they read and how they learned science content. In addition, there are many other questions that we believe can be addressed with this data base. The questions that follow are among those we expect to explore over the next few years.

1. Do teachers teach children of different ability differently when they are at comparable points in a curriculum?

2. What are the short- and long-term effects of different special education treatments?

3. Can we identify a hierarchy of instructional interactions that produce the most effective results in reading/science?

4. How does reading instruction in special classroom settings compare with reading instruction taking place at the same time in regular classrooms? In other words, what are students in special classes getting or missing when they are away from their regular classrooms?

5. Will there be long-term positive effects from teachers' reading to students?

6. How do young children develop reading interests/background knowledge independently?

7. What are the long-term effects of (a) a transition room treatment, (b) holding students over in kindergarten, or (c) holding students over in first grade?

8. Are there long-term effects for whole-day versus half-day kindergarten programs? Or, is there really another issue?

9. Is reading rate as important for older students when they read expository text as it was when they read narrative text in the early grades?

10. What are the differential effects on achievement of silent versus oral reading?

11. What is the relationship between kindergarten children's knowledge of environmental print and their later reading comprehension ability?

12. Are there differences in addition to time spent in reading instruction and reading to students that generally characterize classrooms where teachers' reading to students correlates negatively with students' achievement in reading?

References

Abruscato, J., Fossaceca, J. W., Hassard, J., & Peck, D. (1980). *Holt elementary science*. New York: Holt, Rinehart & Winston.

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79*, 193-223.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the Commission on Reading.* Washington, DC: National Institute of Education.

Anderson, T. H., & Armbruster, B. B. (1984). Content area textbooks. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 195-221). Hillsdale, NJ: Erlbaum.

Barr, R. (1983). *Chicago reading test.* Unpublished instrument. Evanston, IL: National College of Education.

Beck, I. L., & McCaslin, E. S. (1978). *An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs.* Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Beck, I. L., McKeown, M. G., Omanson, R. C., & Pople, M. T. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly, 19*, 263-277.

Bredderman, T. (1983). The effects of activity-based science on student outcomes: A quantitative synthesis. *Review of Educational Research, 53*, 499-518.

Brophy, J. E. (1972). *Stability in teacher effectiveness* (R & D Report Series 77). Austin: The University of Texas, The Research and Development Center for Teacher Education.

Brophy, J. E., & Evertson, C. M. (1976). *Learning from teaching.* Boston: Allyn & Bacon.

Chall, J. S. (1967). *Learning to read: The great debate.* New York: McGraw-Hill.

Chen, Y-M. (1990). Low achieving beginning readers: Early identification and structural modeling of results. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Clymer, J., Wong, O., & Benedict, V. J. (1976). *Reading 720.* Lexington, MA: Ginn and Co. (Xerox Corporation).

College Board (1979). *Degrees of reading power.* New York.

CTB/McGraw-Hill. (1978). *Tests of basic experience.* Monterey Park, CA.

Durr, W. K., LePere, J. M., Alsin, M. L., Burryon, R. P., & Shaw, S. (1979). *Houghton Mifflin reading series.* Boston: Houghton Mifflin Company.

Early, M., Cooper, E. K., & Santeusanio, N. (1983a). *Look, listen, and learn.* Harcourt Brace Jovanovich Bookmark Reading Program, Eagle Edition. New York: Harcourt Brace Jovanovich.

Early, M., Cooper, E. K., & Santeusanio, N. (1983b). *Sounds, symbols, and sense*. Harcourt Brace Jovanovich Bookmark Reading Program, Eagle Edition. New York: Harcourt Brace Jovanovich.

Educational Testing Service. (1976). *CIRCUS reading test*. Menlo Park, CA: Addison-Wesley.

Educational Testing Service. (1979). *Sequential test of educational progress*. Menlo Park, CA: Addison-Wesley.

Engelmann, S. E., & Bruner, E. C. (1983). *Distar reading mastery*. Chicago: Science Research Associates.

Engelmann, S. E., & Meyer, L. A. (1974). *Eugene test of metacognition*. Unpublished instrument. Urbana: University of Illinois, Center for the Study of Reading.

Feitelson, D., Kita, B., & Goldstein, Z. (1986). Effects of listening to stories on first graders' comprehension and use of language. *Research in the Teaching of English, 20*, 339-356.

Hastings, C. N., Meyer, L. A., & Linn, R. L. (1987, April). *Assessing science concept and process acquisition for first-, second-, and third-grade children*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Washington, DC.

Hewison, J., & Tizard, J. (1980). Parental involvement and reading attainment. *British Journal of Educational Psychology, 50*, 209-215.

Holmes, N. J., Leake, J. B., & Shaw, M. W. (1974). *Gateways to science*. New York: Webster Division, McGraw-Hill.

Jarvis, C., & Molnar, J. (1983). *All-day kindergarten program effects on student growth* (O.E.A. Evaluation Report). New York: Central Board of Education.

Jastak, J. F., Bijou, S. W., & Jastak, S. R. (1965). *Wide range achievement test*. Wilmington, DE: Jastak Associates.

Johnson, E. (1974). *An experimental study of the comparison of pupil achievement in the all-day kindergarten and one half-day control group*. Walden University, Naples, FL (ERIC Document Reproduction Service No. ED115 361).

Linn, R. L., & Meyer, L. A. (1985). *Kindergarten instruction and early reading achievement* (Tech. Rep. No. 384). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Linn, R. L., Meyer, L. A., & Hastings, C. N. (1987). *Science instruction in kindergarten and first grade*. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Mallinson, G. G., Mallinson, J. B., Smallwood, W. L., & Valentino, C. (1985). *Silver Burdett science*. Morristown, NJ: Silver Burdett.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. *Child Development, 48*, 986-992.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 50*, 643-655.

McCormick, C., & Mason, J. M. (1986). *Use of little books at home: A minimal intervention strategy that fosters early reading* (Tech. Rep. No. 388). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Greer, E. A., Crummey, L., & Boyer, B. A. (in press). *How do first and second grade reading programs compare?* Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Crummey, L., & Greer, E. A. (1988). Elementary science textbooks: Their contents, text characteristics, and comprehensibility. *Journal of Research in Science Teaching, 25*(6), 435-463.

Meyer, L. A., Hastings, C. N., & Linn, R. L. (1988). *Assessing early reading with new decoding and comprehension measures*. Unpublished paper. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., Linn, R. L., & Greer, E. A. (1985). *The error detection test*. Unpublished manuscript. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., Linn, R. L., & Greer, E. A. (1986). *Description, development, and preliminary findings from a measure of comprehension monitoring for first graders*. Unpublished manuscript. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., & Wardrop, J. L. (1988). *Proposal to the National Science Foundation*. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., Wardrop, J. L., & Linn, R. L. (1988). *How entering ability and instructional settings and not the length of the school day mediate kindergartners' reading performance* (Tech. Rep. No. 466). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Linn, R. L., & Hastings, C. N. (1985a). *A look at instruction in kindergarten: Observations of interactions in three school districts* (Tech. Rep. No. 383). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Linn, R. L., & Hastings, C. N. (1985b). *The heuristic and measurement models during a study of reading comprehension development and science knowledge* (Tech. Rep. No. 382). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Linn, R. L., & Hastings, C. N. (1986). *Teachers' reading to students correlates negatively with student performance in reading: Why might this be?* Paper presented at the annual meeting of the National Reading Conference, Austin, Texas.

Meyer, L. A., Linn, R. L., & Hastings, C. N. (1987). *Parents reports of kindergarten, first- and second-grade children's out of school activities that may be related to science ability and interest* (Tech. Rep. No. 495). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Linn, R. L., & Hastings, C. N. (in press). Teacher stability revisited: How consistent are teachers from morning to afternoon and from year to year? *American Educational Research Journal*.

Nickerson, R. S. (1985). Understanding understanding. *American Journal of Education, 93*, 201-239.

Oliver, L. S. (1980). *The effects of extended instructional time on the readiness for reading of kindergarten children*. Unpublished doctoral dissertation, Boston University.

Osborn, J. (1981). The purposes, uses, and contents of workbooks and some guidelines for teachers and publishers. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 45-110). Hillsdale, NJ: Erlbaum.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart & Winston.

Reiss, E., & Friedman, R. (1976). *Alpha K Time*. Plainview, NY: New Dimensions in Education.

Rogosa, D., Floden, R. E., & Willett, J. B. (1984). *Assessing the stability of teacher behavior* (Research Series No. 141). East Lansing: Michigan State University, The Institute for Research on Teaching.

Rosenshine, B. (1970). The stability of teacher effects upon student achievement. *Review of Educational Research, 40,* 647-662.

Shavelson, R., & Dempsey-Atwood, N. (1976). Generalizability of measures of teaching behavior. *Review of Educational Research, 46,* 553-611.

Stallings, J., Cory, R., Fairweather, J., & Needles, M. (1977). *Early childhood education classroom evaluation*. Menlo Park, CA: SRI International.

Stallings, J. A., & Kaskowitz, D. (1974). *Follow-through observation evaluation, 1972-73.* Menlo Park, CA: Stanford Research Institute.

Sund, R. B., Adams, D. K., Hackett, J. K. (1980). *Accent on science*. Columbus, OH: Charles E. Merrill Publishing Co.

Tizard, J., Schofield, W. N., & Hewison, J. (1982). Collaboration between teachers and parents in assisting children's reading. *British Journal of Educational Psychology, 52,* 1-15.

Weber, G. (1971). *Weber comprehension test*. Unpublished manuscript. Washington, DC.

Winter, M. & Klein, A. (1970). *Extending the kindergarten day: Does it make a difference in the achievement of educationally advantaged and disadvantaged pupils?* Washington, DC: Bureau of Elementary and Secondary Education (ERIC Document Reproduction Service No. ED087 534).

Woodcock, R. W. (1973). *Woodcock reading mastery tests*. Circle Pines, MN: American Guidance Service.

Table 1

Time for: Reading, Instructional Time Minus Story Reading; and Story Reading Aggregated by Grade Level for Kindergarten and First-Grade Teachers (N = 42 classes)

		Time Spent in:							
		Reading Instruction		Instruction Minus Adult Reading		Story Reading		Decoding	
	Grade Level	*x*	*SD*	*x*	*SD*	*x*	*SD*	*x*	*SD*
District A									
(N = 12 classes)	K (Co 1)	40.61	(5.26)	17.89	(1.59)	5.67	(1.46)	122.53	(7.03)
	K (Co 2)	34.97	(3.60)	37.94	(10.47)	4.97	(1.42)	129.67	(10.34)
	1	43.80	(15.41)	108.50	(22.00)	5.89	(4.56)	216.03	(173.28)
District B									
(N = 20 classes)	K (Co 1)	7.57	(6.62)	38.79	(13.54)	7.02	(3.10)	55.87	(17.58)
	K (Co 2)	6.72	(2.22)	38.68	(4.96)	5.20	(3.11)	82.91	(19.93)
	1	38.27	(5.91)	109.40	(14.65)	12.28	(4.97)	254.47	(66.01)
District C									
(N = 10 classes)	K (Co 1)	5.18	(1.99)	23.78	(8.64)	25.56	(9.63)	29.89	(7.58)
	K (Co 2)	5.22	(2.22)	71.44	(20.93)	19.70	(7.21)	20.04	(13.00)
	1	35.20	(12.47)	106.59	(13.57)	4.45	(3.53)	109.30	(39.29)

Table 2

Summary of First- and Second-Grade Basal Reader Decoding Instruction and Text Analysis

							Reading Number of:		
Program	Sounds	Letter Naming	Syllabications/ Endings	Rules	Rhyming	Blending	Vocabulary Words	Words in Stories	Words in Isolation
Ginn, 1976									
1st Grade	2,641	346	0	35	44	0	871	20,982	
2nd Grade	1,502	169	481	35	10	0	1,311	104,600	1,334
Harcourt Brace Jovanovich, 1983									
1st Grade	1,764	670	0	10	292	0	425	17,164	
2nd Grade	3,276	444	1,110	27	161	0	678	79,867	2,658
Houghton Mifflin, 1979									
1st Grade	1,478	808	0	8	119	0	607	12,264	
2nd Grade	3,277	664	567	163	0	0	676	64,669	3,795
SRA Distar Reading Mastery, 1983									
1st Grade	2,655	0	0	0	278	805	1,236	5,919	
2nd Grade	1,339	4,015	0	0	0	579	4,154	67,400	5,702

Table 3

Summary of First- and Second-Grade Basal Reader Comprehension Instruction

	Text Explicit Questions				Text Implicit Questions			
Program	Word	Sentence	Paragraph	Picture	Word	Sentence	Paragraph	Picture
Ginn, 1976								
1st Grade	178	751	0	109	567	1,107	0	472
2nd Grade	1,322	536	103	34	1,951	1,341	219	47
Harcourt Brace Jovanovich, 1983								
1st Grade	216	977	0	232	494	494	0	126
2nd Grade	855	531	166	114	506	1,579	262	50
Houghton Mifflin, 1979								
1st Grade	120	1,192	0	425	851	854	0	273
2nd Grade	274	1,674	111	250	445	506	79	127
SRA Distar Reading Mastery, 1983								
1st Grade	155	541	0	0	0	141	0	0
2nd Grade	38	824	25	384	2	72	12	238

Table 4

Words per Incoherence for Matched and Unmatched First- and Second-Grade Basal Stories

Publisher	Matched *X*	Stories *(SD)*	Unmatched *X*	Stories *(SD)*
Ginn				
1st Grade	27.40	(18.62)	49.38	(29.68)
2nd Grade	62.20	(27.63)	49.09	(19.77)
Harcourt, Brace, Jovanovich				
1st Grade	12.80	(2.37)	27.00	(23.99)
2nd Grade	21.30	(8.73)	15.20	(5.09)
Houghton Mifflin				
1st Grade	78.10	(58.91)	118.50	(29.24)
2nd Grade	43.62	(25.99)	37.82	(12.16)
SRA Distar Reading Mastery				
1st Grade	244.50	(4.95)	104.97	(32.99)
2nd Grade	57.83	(3.38)	68.27	(20.76)

Table 5

Second Graders' Performance in the Content Domains of Plants, Three Forms of Matter, Motion, and on the STEP Science Test

Instrument	District	*X*	*(SD)*
Plants Content Domain Mix of In and Out of Level Items 0 - 33 Score Range	A (*N* = 78) B (*N* = 144) C (*N* = 83)	20.73 22.50 21.22	(3.54) (2.88) (4.51)
Three Forms of Matter Content Domain Out of Level Area 0 - 34 Score Range	A (*N* = 77) B (*N* = 144) C (*N* = 85)	18.95 21.27 20.02	(5.09) (4.61) (6.06)
Motion Content Domain Out of Level Area 0 - 20 Score Range	A (*N* = 77) B (*N* = 144) C (*N* = 85)	11.34 11.87 11.62	(2.66) (2.10) (2.41)
Step Science Test Norm-referenced General Measure of Science Ability 0 - 50 Score Range	A (*N* = 77) B (*N* = 143) C (*N* = 85)	35.86 37.24 35.12	(6.13) (6.93) (8.87)

Table 6

Inconsiderate Structural, Content, and Pictorial Characteristics of Common Content Domains and Grade Levels for the Holt, McGraw-Hill, Merrill, and Silver-Burdett Science Programs

			Number of								
			Structure				Content			Pictures & Diagrams	
Grade Level	Content Domains	Publisher	Illogical Structure	Lack of Connectives or Unclear Referents	Illogical Sequences, Explanations, or Procedures	Irrelevant Ideas	Incomplete Background Knowledge	Problematic Technical Terms	Unnecessary Figurative Language	False Information	Unnecessary	Hard to See or Unclear
1	ANIMALS	Merrill	0	1	0	0	2	0	0	0	0	29
1		Silver-Burdett	5	0	0	0	5	0	0	0	0	54
2		Merrill	2	0	0	1	9	0	0	1	0	27
2		Silver-Burdett	0	0	0	1	14	2	0	0	0	12
4		Merrill	0	6	0	3	28	5	1	0	8	73
4	ELECTRICITY & MAGNETISM	Holt	0	5	0	10	23	1	0	2	5	35
4		Merrill	1	2	0	0	14	0	0	0	0	0
4		Silver-Burdett	0	6	0	3	12	2	0	0	5	11

Table 6 (Continued)

			Number of								
			Structure				Content			Pictures & Diagrams	
Grade Level	Content Domains	Publisher	Illogical Structure	Lack of Connectives or Unclear Referents	Illogical Sequences, Explanations, or Procedures	Irrelevant Ideas	Incomplete Background Knowledge	Problematic Technical Terms	Unnecessary Figurative Language	False Information	Unnecessary	Hard to See or Unclear
4	HUMAN BODY	Merrill	1	6	1	2	14	1	2	1	8	30
4		Silver-Burdett	2	0	0	1	1	0	0	0	4	1
4		Holt	0	0	0	2	3	1	0	1	10	4
5		McGraw-Hill	3	3	0	0	5	0	0	0	4	2
5		Merrill	0	0	0	4	8	5	0	0	9	3
5		Silver-Burdett	0	1	0	2	5	0	0	1	5	4
1	PLANTS	Merrill	5	8	0	6	6	0	0	1	2	6
1		Silver-Burdett	0	0	0	1	8	0	0	0	0	7
2		Merrill	0	11	0	0	11	0	0	0	5	8
2		Silver-Burdett	0	6	0	6	22	0	0	0	0	11

Table 6 (Continued)

			Number of								
			Structure				Content			Pictures & Diagrams	
Grade Level	Content Domains	Publisher	Illogical Structure	Lack of Connectives or Unclear Referents	Illogical Sequences, Explanations, or Procedures	Irrelevant Ideas	Incomplete Background Knowledge	Problematic Technical Terms	Unnecessary Figurative Language	False Information	Unnecessary	Hard to See or Unclear
3	PLANTS (Cont.)	Merrill	4	9	0	6	15	0	0	0	9	10
3		Silver-Burdett	0	1	1	1	10	0	0	0	0	0
4		Merrill	1	6	0	13	18	0	0	0	10	56
4	SOLAR SYSTEM	Holt	0	1	0	1	7	4	1	0	0	8
4		Merrill	0	0	0	0	2	2	0	0	1	25
4		Silver-Burdett	0	4	0	2	12	2	0	0	1	9
5	WEATHER	McGraw-Hill	3	10	0	8	10	0	1	0	4	8
5		Merrill	1	3	0	6	20	2	1	1	0	1
5		Silver-Burdett	0	11	0	3	4	0	0	0	6	12

Table 7

Cross-Publisher Comparisons of Inconsiderate Text Structure and Content in Common Domains by Grade Level

			Number of								
			Structural Problems			Content Problems				Problems with Pictures & Diagrams	
Grade Level	Content Domains	Chapters	Illogical Structure	Lack of Connectives or Unclear Referents	Illogical Sequences, Explanations, or Procedures	Irrelevant Ideas	Incomplete Background Knowledge	Problematic Technical Terms	Unnecessary Figurative Language	False Information	Unnecessary	Hard to See or Unclear
MERRILL, 1980												
1	PLANTS; ANIMALS	4	5	9	0	6	8	0	0	1	2	35
2	PLANTS; ANIMALS	2	2	11	0	1	20	0	0	1	5	35
3	PLANTS	2	4	9	0	6	15	0	0	0	9	10
4	ELECTRICITY & MAGNETISM; HUMAN BODY; SOLAR SYSTEM	7	2	8	1	2	30	3	2	1	9	55
5	HUMAN BODY; WEATHER	6	1	3	0	10	28	7	1	1	9	4
		4.2 (2.3)	2.8 (1.6)	8 (3)	.2 (.44)	5 (3.6)	20.2 (9.1)	2 (3.1)	.6 (.9)	.8 (.4)	6.8 (3.2)	27.8 (20.8)

Table 7 (Continued)

		Number of										
			Structural Problems				Content Problems				Problems with Pictures & Diagrams	
Grade Level	Content Domains	Chapters	Illogical Structure	Lack of Connectives or Unclear Referents	Illogical Sequences, Explanations, or Procedures	Irrelevant Ideas	Incomplete Background Knowledge	Problematic Technical Terms	Unnecessary Figurative Language	False Information	Unnecessary	Hard to See or Unclear
SILVER BURDETT, 1984												
1	PLANTS; ANIMALS	4	0	0	0	2	22	2	0	0	0	19
2	PLANTS; ANIMALS	2	0	6	0	7	36	2	0	0	0	23
3	PLANTS	2	0	1	1	1	10	0	0	0	0	0
4	ELECTRICITY & MAGNETISM; HUMAN BODY; SOLAR SYSTEM	7	2	10	0	6	25	4	0	0	10	21
5	HUMAN BODY; WEATHER	6	0	12	0	5	9	0	0	1	11	16
		4.2 (2.3)	.4 (.9)	5.8 (5.3)	.2 (.4)	4.2 (2.6)	20.4 (11.2)	1.6 (1.7)	0	.2 (.4)	4.2 (5.8)	15.8 (9.2)
HOLT, 1980												
4	ELECTRICITY & MAGNETISM; HUMAN BODY; SOLAR SYSTEM	13	0	6	0	13	33	6	1	3	21	47
MCGRAW-HILL, 1974												
5	HUMAN BODY; WEATHER	2	6	13	0	8	15	0	1	0	8	10

Figure Captions

Figure 1. Heuristic Model.

Figure 2. Annual Data Collection for the Heuristic Model.

Figure 3. Plots of Two District A Teachers' Minutes Allocated to Decoding.

Figure 4. Plots of Three District B Teachers' Minutes Allocated to Decoding.

Figure 5. Plots of Two District A Teachers' Frequencies During Decoding.

Figure 1
Heuristic Model of Reading Comprehension and Science Knowledge Development

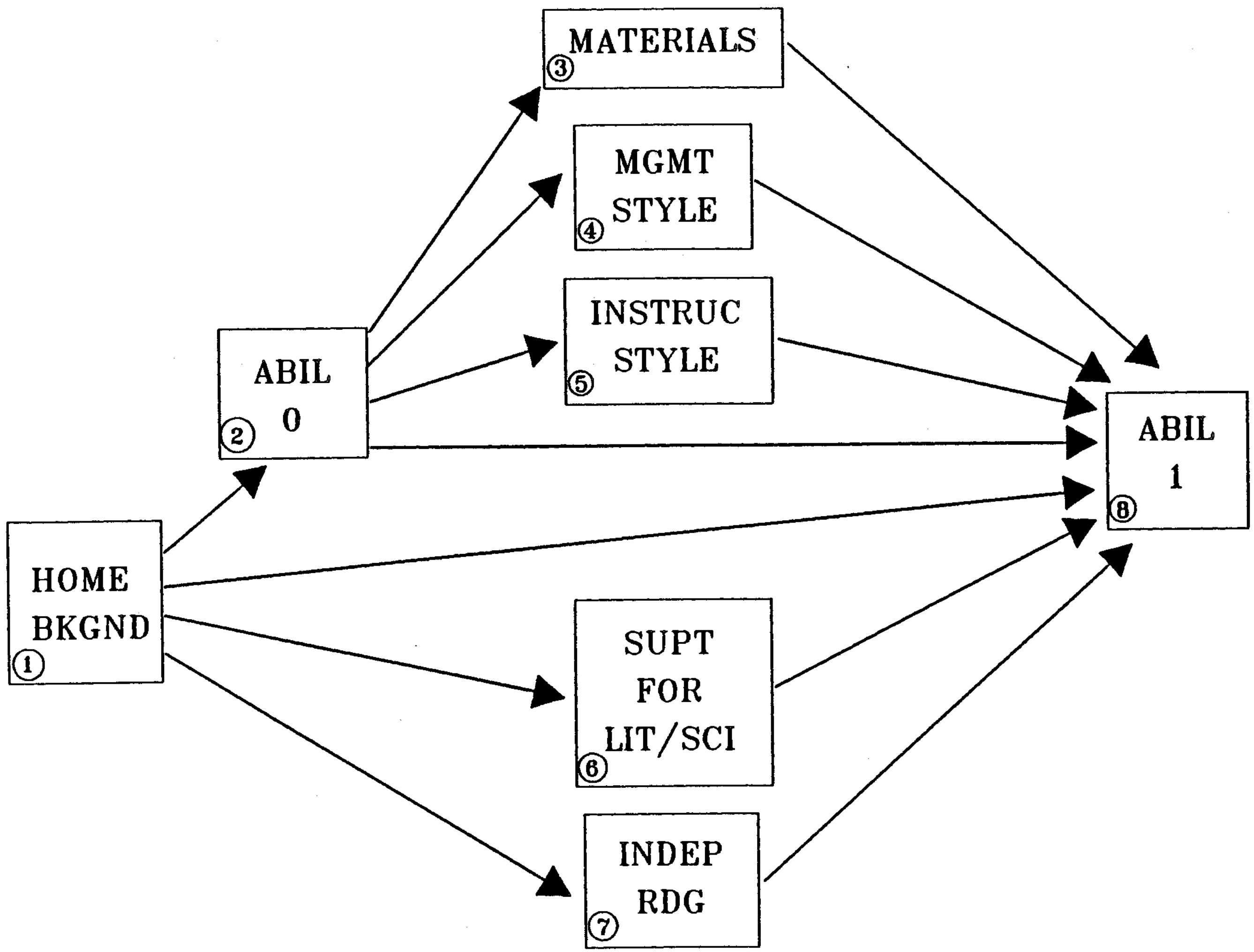


Figure 2

Annual Data Collection for the Heuristic Model

Student Ability
Time 0
(Fall)
Multiple Measures
Norm Referenced
Criterion Referenced
Individual and Group
Assessment
(Many new measures)

Home Background
Parental Occupation
Parental Education
Hours Parents Work
Number of Siblings

Classroom Management
(9 Full Day Observations,
Audio Taped)
1. Time in Activities
2. Organization
(Whole class/Groups)
3. Praise/Corrective
Statements

Materials
(Reading and Science
Programs Every Page)
1. What is there?
2. How is it presented?
3. How "considerate" is it?

Independent Reading

Student Reading
Interest
Opinions

Instructional Style

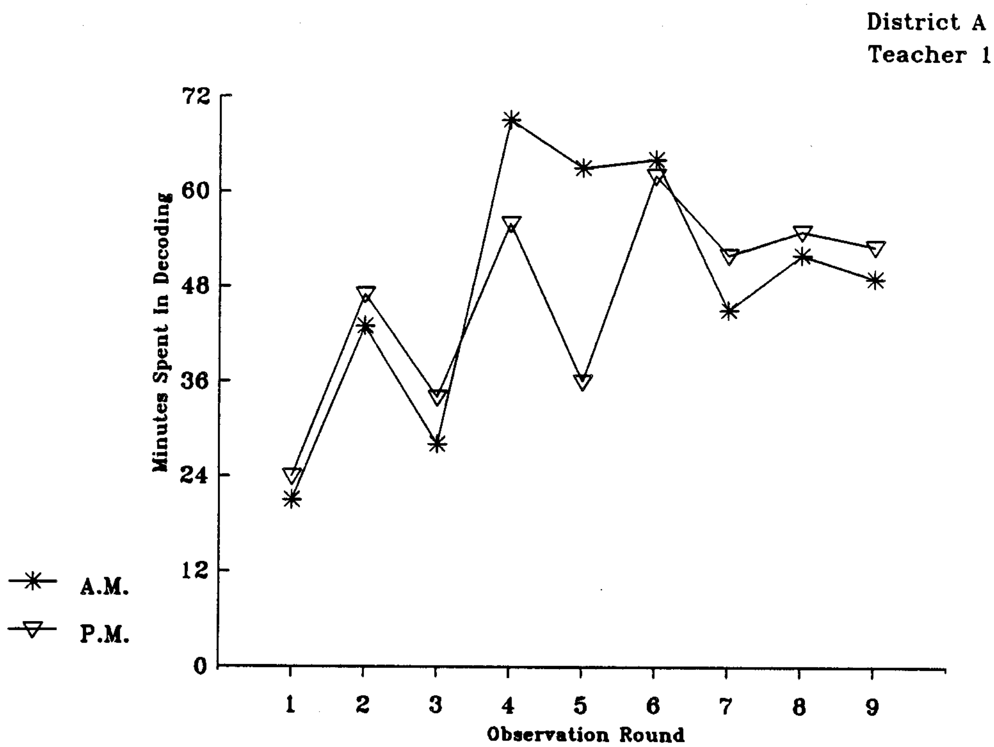
Teachers' Presentations
(9 Full Day Observations,
Audio Taped)
1. What kinds of interactions?
   a. Procedural
   b. Text-Tied Comprehension
   c. Scriptal
   d. Decoding
   e. Word, sentence,
      paragraph reading
   f. Inferential Reasoning
2. How are turns distributed?
   a. Individuals
   b. Small groups
   c. Whole class
3. What kind of feedback?
   a. Sustaining
   b. Terminal
   c. Confirming
4. Seatwork
   a. Type
b. Quantity
   c. Student Performance

Home Influences
(Spring)
Support for Literacy
Support for General Science
Learning

Student Ability
Time 1
(Spring)
Multiple Measures
Norm Referenced
Criterion Referenced
Individual and Group
Assessment
(Many new measures)

Figure 3

Plots of Two District A Teachers' Minutes Allocated to Decoding



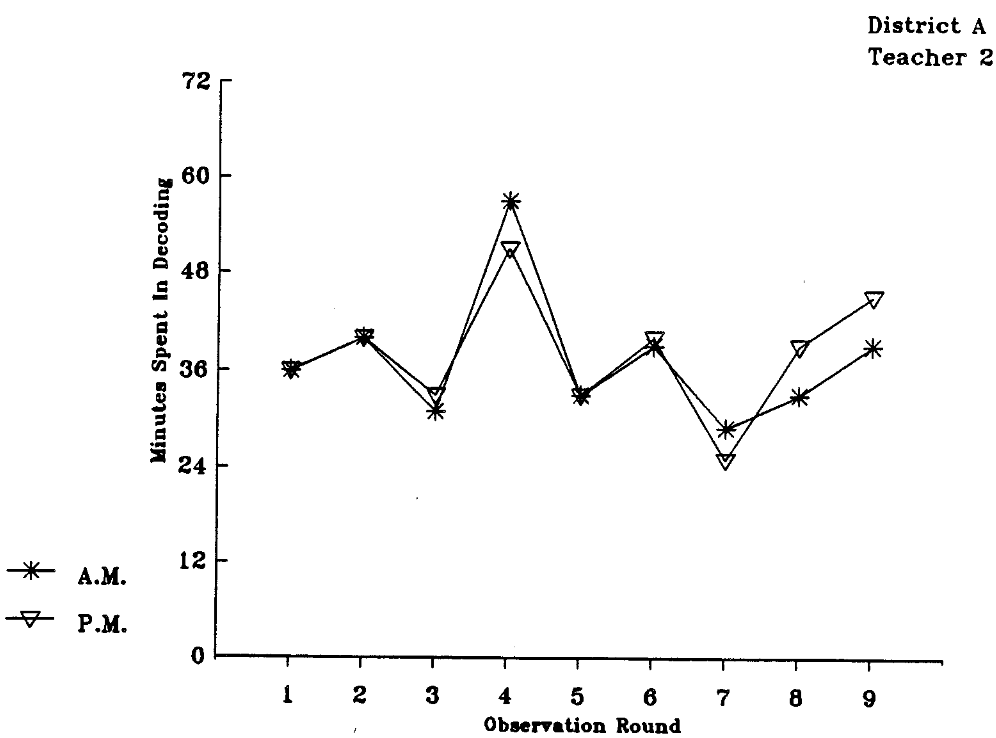
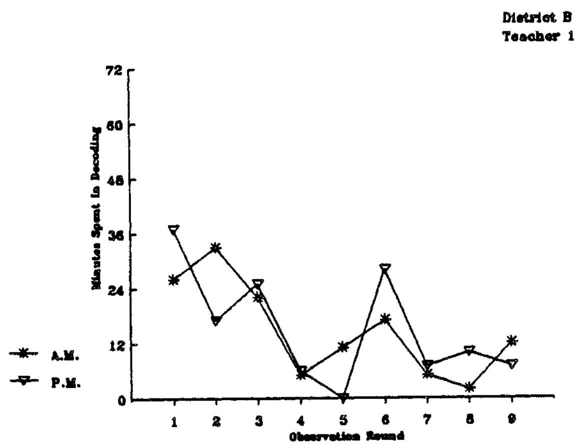
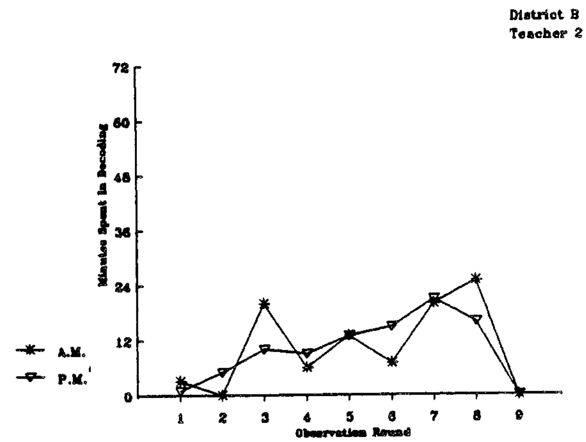


Figure 4

Plots of Three District B Teachers' Minutes Allocated to Decoding





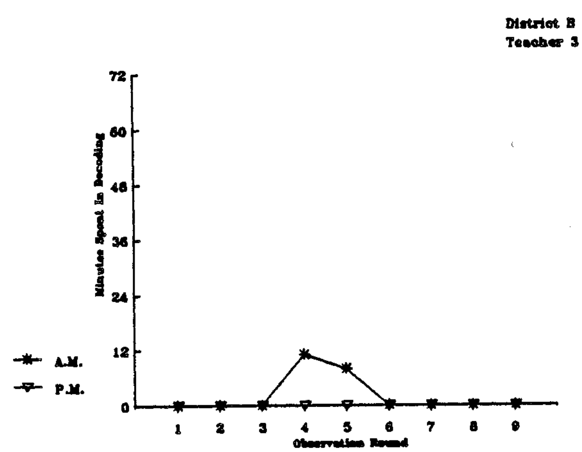
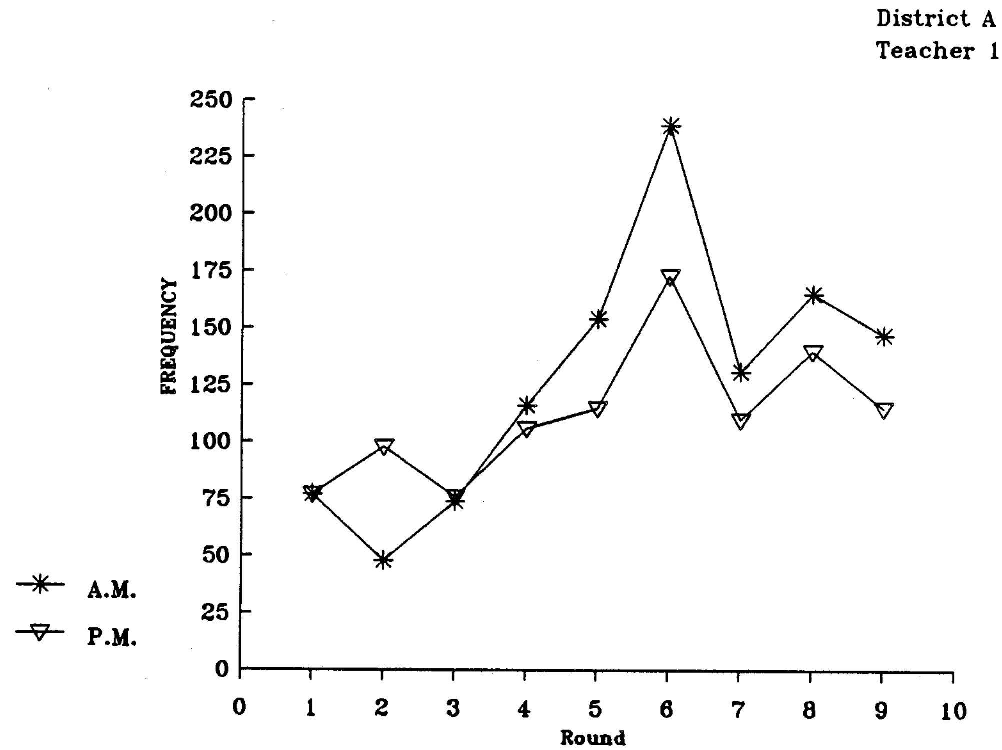
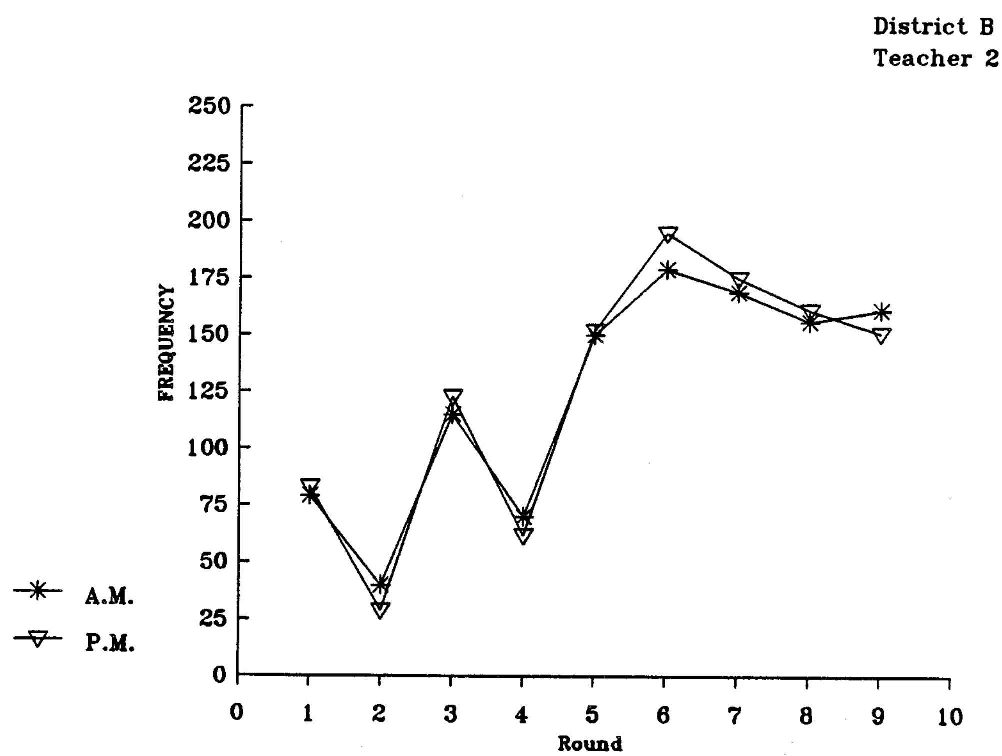


Figure 5

Plots of Two District A Teachers' Frequencies During Decoding





This page is intentionally blank.